Hypothesis Testing

Hypothesis testing is a formal procedure for investigating our ideas about the world using statistics. It is most often used by scientists to test specific predictions, called hypotheses, that arise from theories.

Hypothesis Testing Steps

1. State your research hypothesis as a null hypothesis and alternate hypothesis (H_0) and (H_1).
2. Collect data in a way designed to test the hypothesis.
3. Perform an appropriate statistical test.
4. Decide whether to reject or fail to reject your null hypothesis.
5. Present the findings in your results and discussion section.

Step 1: State your null and alternate hypothesis

After developing your initial research hypothesis (the prediction that you want to investigate), it is important to restate it as a null (H_0) and alternate (H_1) hypothesis so that you can test it mathematically.

The alternate hypothesis is usually your initial hypothesis that predicts a relationship between variables. The null hypothesis is a prediction of no relationship between the variables you are interested in.

Example

You want to test whether there is a relationship between gender and height. Based on your knowledge of human physiology, you formulate a hypothesis that men are, on average, taller than women. To test this hypothesis, you restate it as:

- H_0: Men are, on average, not taller than women.
- H_1: Men are, on average, taller than women.

Step 2: Collect data

For a statistical test to be valid, it is important to perform sampling and collect data in a way that is designed to test your hypothesis. If your data are not representative, then you cannot make statistical inferences about the population you are interested in.

Example

To test differences in average height between men and women, your sample should have an equal proportion of men and women, and cover a variety of socio-economic classes and any other control variables that might influence average height.

You should also consider your scope (Worldwide? For one country?) A potential data source in this case might be census data, since it includes data from a variety of regions and social classes and is available for many countries around the world.

Step 3: Perform a statistical test

There are a variety of statistical tests available, but they are all based on the comparison of within-group variance (how spread out the data is within a category) versus between-group variance (how different the categories are from one another).

If the between-group variance is large enough that there is little or no overlap between groups, then your statistical test will reflect that by showing a low p-value. This means it is unlikely that the differences between these groups came about by chance.

Alternatively, if there is high within-group variance and low between-group variance, then your statistical test will reflect that with a high p-value. This means it is likely that any difference you measure between groups is due to chance.

Your choice of statistical test will be based on the type of variables and the level of measurement of your collected data.

Example

Based on the type of data you collected, you perform a one-tailed t-test to test whether men are in fact taller than women. This test gives you:

an estimate of the difference in average height between the two groups.
a p-value showing how likely you are to see this difference if the null hypothesis of no difference is true.

Your t-test shows an average height of 175.4 cm for men and an average height of 161.7 cm for women, with an estimate of the true difference ranging from 10.2 cm to infinity. The p-value is 0.002.

Step 4: Decide whether to reject or fail to reject your null hypothesis

Based on the outcome of your statistical test, you will have to decide whether to reject or fail to reject your null hypothesis.

In most cases you will use the p-value generated by your statistical test to guide your decision. And in most cases, your predetermined level of significance for rejecting the null hypothesis will be 0.05 – that is, when there is a less than 5% chance that you would see these results if the null hypothesis were true.

In some cases, researchers choose a more conservative level of significance, such as 0.01 (1%). This minimizes the risk of incorrectly rejecting the null hypothesis (Type I error).

Example

In your analysis of the difference in average height between men and women, you find that the p-value of 0.002 is below your cutoff of 0.05, so you decide to reject your null hypothesis of no difference.